

VIMS/Bioinformatics HPC Tutorial

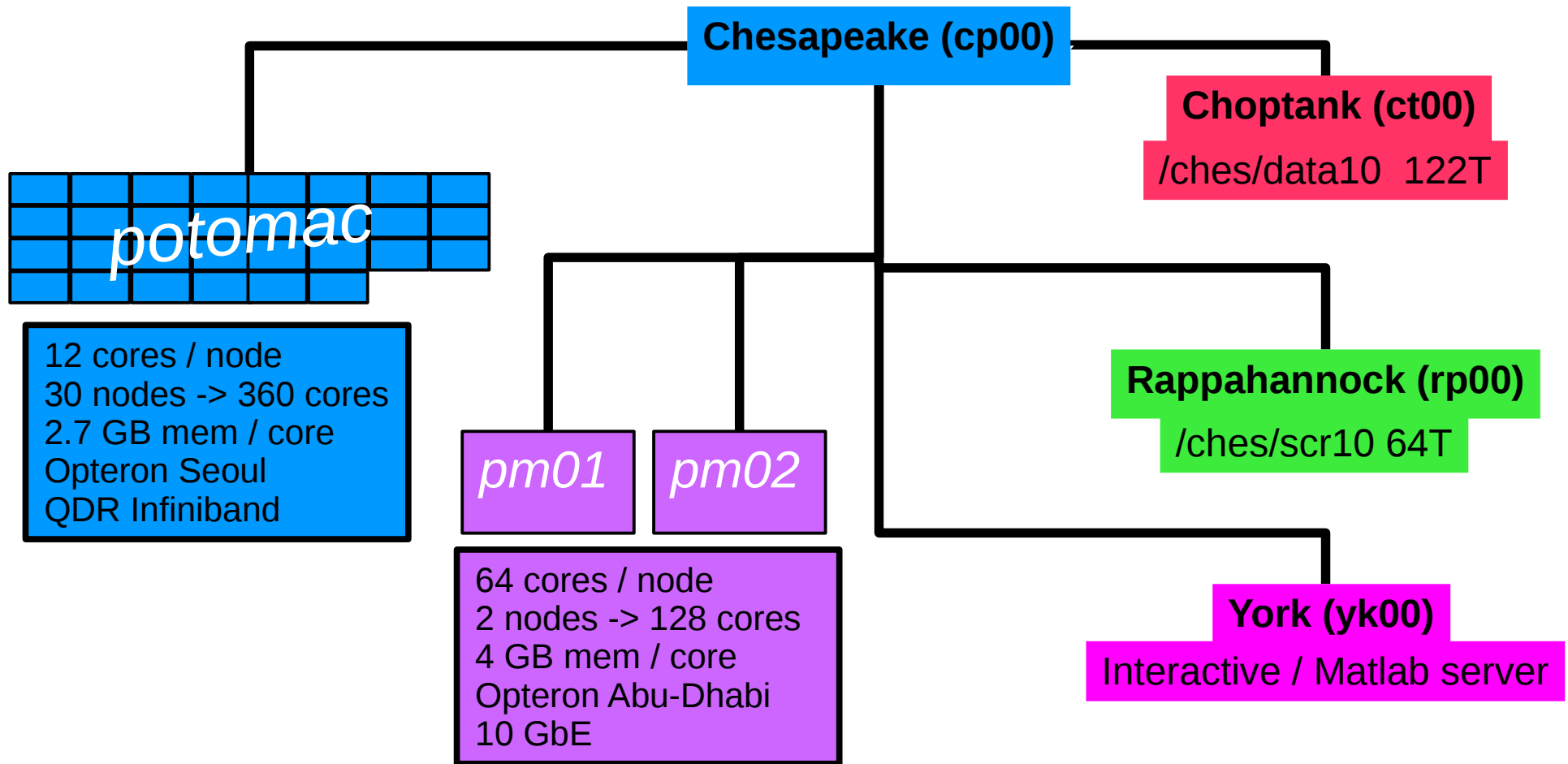
- Understanding the available hardware and software
- Using software modules
- Overview of file systems
- How to use the batch system to run jobs

March 5th, 2018

**Eric J. Walter
Manager of HPC
W&M and VIMS**

Chesapeake

.hpc.vims.edu



- **Chesapeake (cp00)** – front-end / login server
- **Potomac** – MPI / multinode parallel
- **Pamunkey** – “Bio” nodes (shared memory parallel)
- **York (yk00)** – Interactive / Matlab server
- **Choptank (ct00) / Rappahannock (rp00)** – file servers

Bioinformatics Software on PM

program	version	program	version	program	version
archaeopteryx	---	jellyfish	2.2.6	quast	v5.0.0
biopython	1.66	JProfileGrid	v2.0.5	rainbow	2.0.4
bedtools	v2.27.1	last-align	---	rdp-classifier	2.11
blast2	2.2.25	lastz	1.02.00	readseq	2.1.30
bowtie	1.1.2	mesquite	3.1.0	samtools	1.3.1
BTLib	0.19	metaxa2	2.1	SPAdes	v3.11.1
bwa	0.7.15-r1142	metaphlan2	2.5.0	structure	2.3.4
clearcut	1.0.9	metavelvetg	1.2.02	scons	2.5.0
clustalw	2.1	mira-3rdparty	4.0.2	trimmomatic	0.36
cufflinks	v2.2.1	mira-assembler	4.0.2	seqprep	---
dDocent	2.2.15	mothur	v1.38.1	sickle	1.33
dendroscope	3.5.7	mothur-lookup-files	---	sortmerna	2.1b
emirge	---	multiqc	1.4	splitstree4	---
fasta	36.3.8d	mummer	3.23	ssake	v.3.8.4
fastqc	v0.11.5	muscle	v3.8.31	usearch	v9.1.13
fastStructure	---	mview	1.6	velvet	v1.2.10
fasttree	2.1.9	ncbi-blast+	0.2.5	velvet-long	v1.2.10
fastx-toolkit	---	ncbi-blast+-legacy	---	vsearch	v2.3.4
freebayes	v1.0.2	phylml	20120412	wgsim	0.3.1-r13
hmmer	3.1b2	picard-tools	---	Xander_assembler	---

Necessary Topics for HPC

- How to log into HPC machines?
- Linux Shell / Text editors – basic Linux skills.
- What software do I want to run? (do I need to compile?)
- What sub-cluster will I use?
- What file-system should I use?
- Using the batch system.

Where to get help?

HPC webpage:	https://www.wm.edu/offices/it/services/hpc/atwm/index.php https://hpc.wm.edu
HPC ticket system	mail: hpc-help@wm.edu

Logging into HPC machines

Must use Secure Shell client (SSH)

- Linux / Mac built-in (terminal)
- Windows – SSH Secure Shell Client / PuTTY

```
11 [chinook] ssh ches.hpc.vims.edu
```

```
Password:
```

```
Last login: Sat Mar 3 19:18:16 2018 from chinook.sciclone.wm.edu
```

```
-----  
Chesapeake Cluster
```

```
College of William & Mary / Virginia Institute of Marine Science
```

```
Info & Documentation: http://www.wm.edu/offices/it/services/hpc/using
```

```
Assistance & Trouble Reports: hpc-help@lists.wm.edu https://hpc.wm.edu/support  
-----
```

- **Am I on or off campus?**

If you are off-campus

- log into *stat.wm.edu* first using your **W&M username and password**
- use VIMS VPN service

- **Is my username the same as my current machine?**

If it is different use: `ssh <username>@<host>.<domain>`

- **Do I need graphics?**

If yes, then log in with `-X`

Linux Shell Usage

Main things to learn about linux/shell

- **learn to log in and out of front-end servers**

ssh, X forwarding, alternate user name

- **manipulating files/folders**

cd – change directory

cp – copy file

mv – rename file

man – read manual page

rm – remove file

mkdir – make directory

rmdir – remove directory

- **configuring your environment**

env – print your environmentals variables

editing your .cshrc.\$PLATFORM file

Linux Text Editors

Popular text editors: emacs or vi/vim

emacs: huge, bloated, not installed by default, but the champion!

vi/vim: tiny, always available, some users love it (?)

nano: editor with training wheels, very easy to use, not very powerful

Emacs Tutorials

<http://www.gnu.org/software/emacs/tour/>

<http://www.jesshamrick.com/2012/09/10/absolute-beginners-guide-to-emacs/>

<http://www2.lib.uchicago.edu/keith/tcl-course/emacs-tutorial.html>

Vim Tutorials

<http://www.vim.org/>

<http://vim.wikia.com/wiki/Tutorial>

<http://linuxconfig.org/vim-tutorial>

Nano homepage

<http://www.nano-editor.org/>

HPC Filesystems / Backup

```
11 [chesapeake] df -h
Filesystem                Size      Used Avail Use% Mounted on
/dev/mapper/VG00-LV00    2.0G      1.2G   736M  62% /
tmpfs                    16G         16K    16G   1% /dev/shm
/dev/sda1                194M      33M   152M  18% /boot
/dev/mapper/VG02-LV01    551G      301G   222G  58% /ches/home00
/dev/mapper/VG00-LV07    43G       8.4G    32G  22% /ches/scr00
/dev/mapper/VG00-LV05    50G       31G    17G  65% /export
/dev/mapper/VG00-LV02    7.9G      189M    7.3G   3% /tmp
/dev/mapper/VG00-LV03    9.9G      6.1G    3.4G  65% /usr
/dev/mapper/VG00-LV06   148G     101G    40G  72% /usr/local
/dev/mapper/VG00-LV04    7.9G      3.6G    3.9G  48% /var
yk00-ib:/york/scr10     1.1T      201M    1.1T   1% /york/scr10
ct00-ib:/ches/data10   122T     110T    13T   90% /ches/data10
rp00-ib:/ches/scr10    61T      14T     48T  23% /ches/scr10
```

home00 – backed up daily – use for small data sets / text files / code etc.

data10 – backed up weekly – long term storage of **current** projects

scr10 – not backed up – used for run-time I/O for calculations

Environment / Start up scripts

<http://www.wm.edu/offices/it/services/hpc/using/shell/index.php> - environment help

<http://www.wm.edu/offices/it/services/hpc/using/modules/index.php> - module help

All software is controlled by software modules

All startup software modules are controlled by ***.cshrc.\$PLATFORM***

The ***\$PLATFORM*** variable is defined on all servers and nodes:

potomac / york : ***\$PLATFORM = rhel6-opteron***

pamunkey : ***\$PLATFORM = rhel7-opteron***

To set startup modules for pm nodes, edit ***.cshrc.rhel7-opteron***

To load correct software path (correct architecture), must load correct ***isa module***

potomac : ***isa/seoul***

pamunkey : ***isa/abu-dhabi***

york : ***isa/magny-cours***

This is also import for compiling for a particular target.

Torque Batch System I

<http://www.wm.edu/offices/it/services/hpc/using/jobs/index.php> - jobs help

Batch system controls all job execution. **Should not run jobs outside of batch system**
Exceptions are short / small memory runs on front-ends

Two types of jobs: interactive and batch

```
20 [chesapeake] qsub -I -l walltime=20:00 -l nodes=1:pamunkey:ppn=1
qsub: waiting for job 97061.cp00.hpc.vims.edu to start
qsub: job 97061.cp00.hpc.vims.edu ready
```

```
11 [pm01]
```

Can also run GUI from interactive job if X is forwarded:

```
21 [chesapeake] qsub -I -l walltime=20:00 -l nodes=1:c19:ppn=1 -X
qsub: waiting for job 97061.cp00.hpc.vims.edu to start
qsub: job 97062.cp00.hpc.vims.edu ready
```

```
11 [pm01]
```

To end interactive job, logout:

```
11 [pm01] logout
```

```
qsub: job 97062.cp00.hpc.vims.edu completed
22 [chesapeake]
```

Torque Batch System II

<http://www.wm.edu/offices/it/services/hpc/using/jobs/index.php> - jobs help

“Batch” jobs are run non-interactively. Controlled with a batch script:

```
22 [chesapeake] cat run

#!/bin/tcsh
#PBS -N test
#PBS -l walltime=1:00:00
#PBS -l nodes=1:pamunkey:ppn=4
#PBS -j oe

cd $PBS_O_WORKDIR

setenv WORKDIR /ches/scr00/ewalter/vims-tool-test-master/data
cd $WORKDIR
mkdir fastqc_results
fastqc *.fastq -t 4 -o fastqc_results |& tee
$WORKDIR/fastqc_stdout.step1

echo "Step 1 - FastQC Completed"
```

To launch job:

```
23 [chesapeake] qsub run
97062.cp00.hpc.vims.edu

24 [chesapeake]
```

Torque Batch System III

<http://www.wm.edu/offices/it/services/hpc/using/jobs/index.php> - jobs help

Once the job runs, batch system returns stdout and stderr files which contain any output for the job in the form of

<JOB-NAME>.o<ID> (and **<JOB-NAME>.e<ID>** if **-j oe** is not in batch script).

qsub – submit a batch or interactive job
qstat – List all jobs in batch system
qsu – List just my jobs
qdel – delete a job

Where to get help?

HPC webpage:	https://www.wm.edu/offices/it/services/hpc/atwm/index.php https://hpc.wm.edu
HPC ticket system	mail: hpc-help@wm.edu